Marine Environmental Data Service

# On the Forecasting of Extreme Sea States

**P.H. LeBlond**
**Seaconsult Marine Research Ltd.**

**CONTRACTOR REPORT NO. 2**

**June 1981**

Government of Canada
Fisheries and Oceans

Gouvernement du Canada
Pêches et Océans

Canada

## MARINE ENVIRONMENTAL DATA SERVICE

CONTRACTOR REPORT NO. 2

# On the Forecasting of Extreme Sea States

P.H. LeBlond, Ph.D.
Seaconsult Marine Research Ltd.
Vancouver, B.C.

DEPARTMENT OF FISHERIES AND OCEANS
OTTAWA, CANADA

© Minister of Supply and Services Canada

This report was prepared by:

Seaconsult Marine Research Ltd., 1001-1111 West Hastings Street, Vancouver, B.C. under DSS Contract OSZ80-00144. The contents of this report are the responsibility of the contractor.

Scientific Authority - J.R. Wilson, MEDS.

Requests for copies should be sent to:

Marine Environmental Data Services Branch, Ocean Science and Surveys, Department of Fisheries and Oceans, Ottawa, Ontario. K1A 0E6.

Correct citation for this publication:

LeBlond, P.H. 1981. On the forecasting of extreme sea states. Mar. Environ. Data Serv. Contract. Rep. Ser. 2:46p.

Ottawa, 1981

MEDS8100702ZE

Since the affairs of men rest still uncertain,
Let's reason with the worst that may befall.

SHAKESPEARE

## ABSTRACT

The principal methods of estimating "design-wave" conditions using various extreme value analysis techniques are examined critically. The uses of measured wave data and wave data hindcast from winds in the process of establishing expected maximum waves in a given period of time are discussed. Special considerations associated with local conditions or wave breaking are also discussed. General conclusions and recommendations on establishing design wave parameters are presented.



## RESUME

Les principales méthodes d'estimation des conditions de vagues "fictives" à l'aide des diverses techniques d'analyse des valeurs extrêmes sont examinées d'un point de vue critique. On étudie l'utilisation des données de mesure des vagues et des rétroprévisions établies à partir des vents, destinée à établir les vagues maximales prévues au cours d'une période de temps donnée. On examine aussi des considérations spéciales relatives aux conditions locales ou au déferlement des vagues. Les conclusions générales et les recommandations sur l'établissement des paramètres de vagues fictives sont présentées.

# Table of Contents


| | Page |
|---|---|
| Table of Contents | i |
| List of Tables | ii |
| List of Figures | iii |
| 1. Introduction | 1 |
| 2. Definition of Wave Quantities | 2 |
| 3. The Available Wave Information | 6 |
| 3.1 Wave Statistics | 6 |
| 3.2 Wind Wave Generation | 10 |
| 3.3 Storm Surges | 13 |
| 4. Methods of Inference | 14 |
| 5. Extrapolation from the Observed Distribution | 16 |
| 6. Extreme Value Analysis | 19 |
| 6.1 Plotting Rules | 20 |
| 6.2 Inferring a Design Wave Height | 24 |
| 6.3 Errors | 27 |
| 6.4 Applications | 29 |
| 7. The Hindcasting Method | 31 |
| 8. Special Considerations | 36 |
| 8.1 Wave Breaking | 36 |
| 8.2 Local Effects | 36 |
| 9. Conclusions and Recommendations | 38 |
| 10. Bibliography | 41 |

## List of Tables


| Table | | Page |
|---|---|---|
| I | Plotting factors u, v for a selection of distributions. | 22 |
| II | Fitting positions and calculated reduced variates for a set of extreme wave height values. | 23 |

## List of Figures


| Figure | | Page |
|---|---|---|
| 1 | Definition of quantities used to describe sea level variations. | 4 |
| 2 | Probability $P\left(\frac{H}{m_o^{\frac{1}{2}}} > x = \frac{H_o}{m_o^{\frac{1}{2}}}\right)$ of exceeding a normalized wave height $x = H_o/m_o^{\frac{1}{2}}$. | 8 |
| 3 | Comparison of maximum measured and model values of significant wave height and zero-crossing period at stations Famita and Stevenson. | 12 |
| 4 | Fits of the lines $x = Ar+B$ to extreme wave height data from Table II. | 25 |
| 5 | Illustrating the influence of sample variability in fitting distributions of maximum wave height E to the reduced variate r. | 26 |
| 6 | The effect on long-term extrapolation of not knowing the "true" distribution. | 28 |
| 7 | The spreading of confidence intervals about a FT-I fit to the extreme wave heights of Table II. | 30 |

## 1. Introduction

Predicting the future is a risky business! Yet, it is an essential step in all of man's planning. From the primitive cycle of planting and harvesting to modern risk analysis, we can characterize the progress of civilization as the evolution of man's skill at avoiding tomorrow's surprises. It is within that broad context of future-gazing, and in parallel with other similar activities that one should place the topic of estimating extreme sea states. The estimation of worst wave conditions is similar to weather prediction or maximum flood calculations and other acts of futurology, whereby information from the past is analyzed and used to construct models which blend deterministic and stochastic ideas and assumptions about the future.

This review of the methods of estimating "design-wave" conditions first defines (in Chapter 2) the quantities of interest; the essential information available on wave statistics and generation is briefly reviewed in Chapter 3 and a general description of the methods of statistical inference used to relate present knowledge to future probabilities is given in Chapter 4.

The principal methods used for design wave estimation are discussed in more detail in subsequent chapters (5 to 7), with particular attention given to matters of applicability, required database and confidence limits. Special considerations associated with local conditions or wave breaking are discussed in Chapter 8. General conclusions and recommendations are presented in the last section (Chapter 9).

## 2. Definition of Wave Quantities

Quantities of interest in a discussion of sea-waves include variations in sea-level, in an absolute sense as measured from the bottom, as well as relative changes about a short-time mean, the water currents associated with these changes and the forces exerted on structures by these currents. Our discussions will focus on sea-level variations: it will be assumed that it is possible, at least in principle, to infer currents and hence forces from a knowledge of those variations, or if the link cannot be completed between wave heights and currents that the latter may be inferred by the same statistical methods as the former.

Variations in sea-level with respect to a fixed reference level may be caused by a variety of phenomena (LeBlond and Mysak, 1978). In order of increasing period, wind waves come first and are our principal concern (periods 1-25 sec); tsunamis, generated by earthquakes or submarine landslides (Murty, 1977) are rare but may be important in shallow water (periods 5-150 min); storm surges are large-scale changes in sea-level occurring in strong storms, often causing extensive flooding on low coasts and with periods of the order of hours to days; tides are mainly of semi-diurnal and diurnal periods and may cause sea-level variations of many meters. Longer period sea-level changes due to seasonal or climatic variations are much less important and will be neglected.

Important components of sea-level variability will include wind waves, tides and storm surges; tsunamis are rare and causally uncorrelated with any of the above phenomena and may be considered as a completely independent risk. Of the main components, the tides are almost fully deterministic and can be forecast far into the future using routine methods (Foreman, 1977). Only wind waves and storm surges are of direct interest here.

At any location, the depth of water D(t) varies rapidly during a storm because of the passage of wind-generated waves and more slowly because of large-scale adjustments of sea-level under atmospheric pressure and wind stress variations. As the periods of wind waves never approach storm time-scales, there is a spectral gap in the frequency domain which allows separation of D(t) into two parts (Figure 1):

$$D(t) = D_o(T) + \zeta(t)$$

Where $\zeta(t)$ represents the wind waves and $D_o(T)$ the storm surges, varying on a longer time scale T, over which the average of the wave contributions vanish: $\langle\zeta(t)\rangle = o$. Statistics of the rapid sea-surface variations $\zeta(t)$ about $D-D_o$ may be obtained for various choices of features. Crest heights are those levels $\zeta_c(t)$ where $\partial\zeta/\partial t = 0$ and $\partial^2\zeta/\partial t^2 < 0$; some values of $\zeta_c$ may be negative (Figure 1). Wave heights are usually defined as the difference between successive low and high (or high and low) values of $\zeta(t)$ separated by a zero-crossing (to eliminate smaller waves, of little interest); these are respectively zero "up-crossing" and zero "down-crossing" wave heights $H_{up}$ and $H_{down}$ (Figure 1).

The wave period may be defined in a number of fashions: between successive crests, or successive troughs, or between successive zero up- or down-crossings. The latter choices ensure that the larger waves are properly described and the period estimates are not affected by the smaller variations of $\zeta(t)$.

Although some of the statistics of the sea surface have been expressed in terms of crest-heights (Cartwright & Longuet-Higgins, 1956), most of the literature on design wave estimation is in terms of zero-crossing wave heights H (we shall make no distinction between up- and down-crossings). An advantage of dealing with wave heights rather than crest heights is that the statistics of the former are nearly independent, for normal sea-states, of spectral bandwidth (ref. Jahns and Wheeler, 1973), i.e. of the

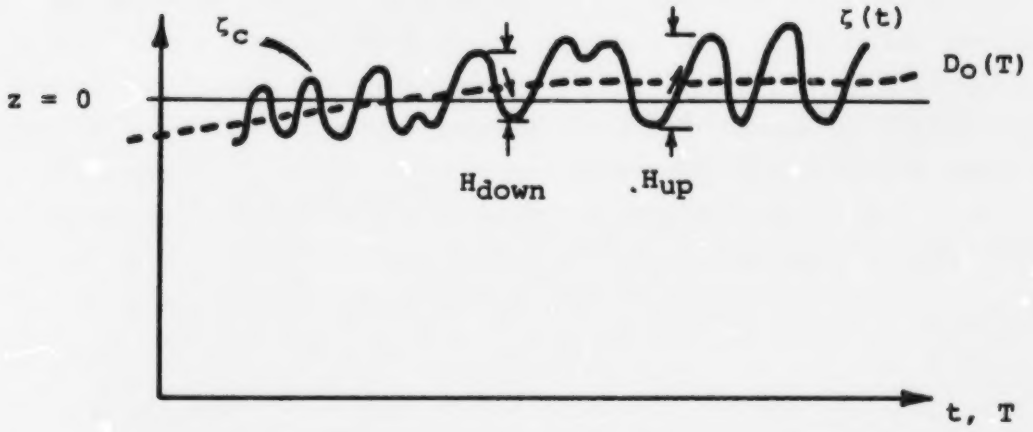


Figure 1 Definition of quantities used to describe sea level variations. $\zeta_c$ = crest height; $H_{down}$, $H_{up}$ = zero down- and up-crossing wave heights; $D_o(T)$ is the sea-level averaged over a 20 min or so sampling period, $\zeta(t)$ is the rapid sea-level variation about $D_o(t)$.

presence of higher frequencies which increase the number of crests without affecting the number of zero crossings. Because of nonlinear effects, however, sea-level variations are not symmetric about the mean: large waves tend to have sharp crests and flat troughs, so that the crest height above mean level is more than half the wave height. This is an important consideration when interpreting design wave height values in terms of absolute water depths. Theoretical results (Longuet-Higgins, 1973) show that for the steepest waves in deep water the ratio of crest height above mean level to wave height from trough to crest reaches 0.68 (it would be equal to 0.5 for a symmetric wave); observational results (Jahns and Wheeler, 1973) give ratios of 0.7 to 0.6 for the largest waves, with the larger values corresponding to higher distortion, occurring in shallower water.

In conjunction with zero-crossing wave heights, the periods used will be the mean interval between zero-up (or down) crossings,

denoted $T_z$. Under conditions of active wave generation, where the wave spectrum is sharply peaked, this measure of period is strongly correlated (Wilson and Baird, 1972) with the period of the spectral peak $T_p$ and with the period $T_{sig}$ (the mean of the periods of the highest one-third of the crest-to-trough waves). In conditions where a large swell is present, or a less-well-developed sea has been reached, so that the spectrum is bimodal, or at least flatter than for a fully developed sea, $T_z$ may be retained as a useful measure of wave period, but it will be less clearly related to $T_p$ or $T_{sig}$.

Questions of primary concern in design wave estimation are:

i) What is the maximum absolute sea level height $D_o(T) + \zeta(t)$ expected during a specified period (50 or 100 years, for example, or the expected working lifetime of a structure)?

ii) What is the maximum zero-crossing wave height $H_{max}$ expected during the same period?

iii) What is the period $T_z$ associated with $H_{max}$?

The answers to these questions are, of course, statistical estimates and subject to confidence estimates. Another way of bringing out the statistical nature of the problem is to rephrase the questions as: "What is the probability that a wave height (or other quantity) will not exceed a certain value over the specified period?"

## 3. The Available Wave Information

Estimates of design wave height must of necessity be based on statistical inferences, either using data already in hand about the statistics of sea-level variations or, at a level once-removed from the actual phenomena, estimating probable future storm intensities and relating them to wave heights and storm-surge levels through a wave-generation and storm-surge model. An even higher level in the causative hierarchy would involve the elaboration of a climate model from which future trends in storm intensity and frequency might be deduced. Although there are indications, at least in the North Sea (Rye, 1976) of significant climatic effects, this latter level of modelling sophistication still seems beyond practical applicability and only wave statistics and the generation of wind waves and storm surges are discussed as part of the available data base.

### 3.1 Wave Statistics

A widely favoured representation of the space-time history of the variations in sea-level is in terms of a Fourier superposition of travelling waves of a wide range of frequencies on a two-dimensional surface. A two-dimensional spectrum $E(\omega, \underset{\sim}{\kappa}, t)$ describes the distribution of sea-surface variance (about the sampling mean) in angular frequency ($\omega$) and wave-number ($\underset{\sim}{\kappa}$) space, and its evolution with time, t. For a narrow spectrum (in both $\omega$ and $\underset{\sim}{\kappa}$) of uncorrelated waves, Longuet-Higgins (1952) showed that the sea-surface displacement should obey random-walk statistics according to which the wave heights would follow the Rayleigh distribution:

$$P(H \geq H_o) = \exp[-H_o^2/8m_o] \qquad (1)$$

where the left hand side denotes the probability that the wave height H exceeds a value $H_o$, and $m_o$ is the variance of the sea-level variation during the sampling period. As a consequence of (1), the average height of the highest one-third of the wave

heights, denoted by $H_{1/3}$ is given by

$$H_{1/3} = 4.005\, m_o^{1/2} \qquad (2)$$

This relation is now routinely used to compute the "significant wave height" $H_{sig} = H_{1/3}$ from an instrumental record, and will be taken as a definition of $H_{sig}$.

Although early comparisons of observed wave heights statistics with the Rayleigh distribution tended to confirm its appropriateness (Goodknight and Russell, 1963; Wilson and Baird, 1972; Borgman, 1973; Earle, 1975), further examination has shown it to overestimate the higher waves in a record by about 10% (Jahns and Wheeler, 1975; Haring et al., 1976; Forristall, 1978; Nolte and Hsu, 1979). A fit with the Weibull distribution

$$P(H \geq H_o) = \exp\left[-\frac{1}{\beta}\left(\frac{H_o}{m_o^{1/2}}\right)^{\alpha}\right] \qquad (3)$$

has been found to give better results. Nolte and Hsu (1979) find $\alpha = 2.138$, $\beta = 8.534$ as the best fit for Hurricane Camille waves. Forristall (1978) obtains $\alpha = 2.126$, $\beta = 8.42$ for waves from 8 sites in the Gulf of Mexico (Figure 2). We note, comparing (3) to (1), that $\alpha = 2$, $\beta = 8$ yields the Rayleigh distribution as a special case. Confidence intervals at the 90% level calculated from Forristall's data give $\Delta\alpha \simeq \pm 0.02$, $\Delta\beta \simeq \pm 0.25$, so that the departures from the Rayleigh distribution are certainly significant at that level.

Discussions of the cause of the failure of the Rayleigh distribution to describe the wave height statistics have failed to pinpoint which precise features of wind waves are responsible for the discrepancy with Longuet-Higgins' (1952) simple model. Spectral width considerations extending the theory to the distribution of crest heights of a broad spectrum (Cartwright and Longuet-Higgins, 1956) are not relevant to the statistics of zero-crossing wave heights. Forristall (1978) and Jahns and Wheeler (1975) speculate that the skewness of the wave profile distribution associated with nonlinearity, whereby crests are higher than troughs are

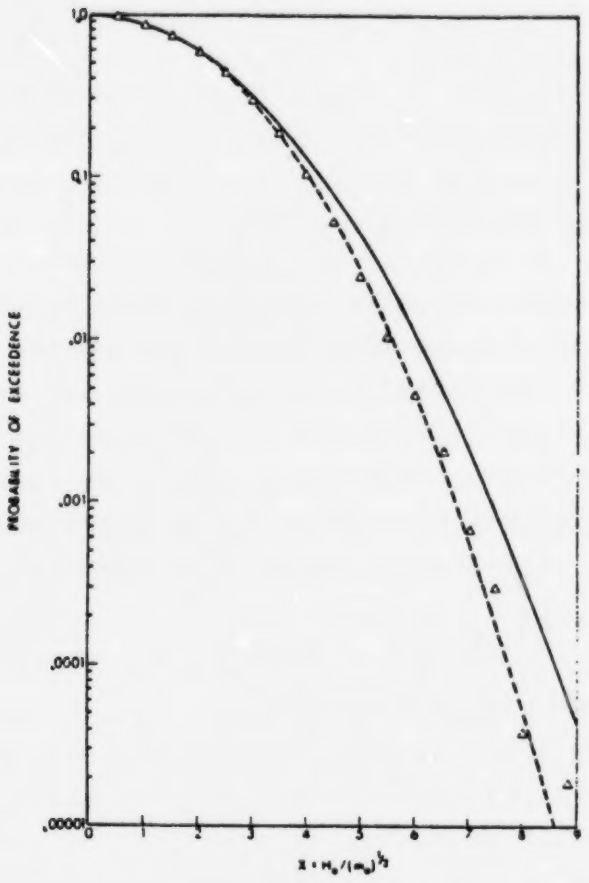


Figure 2 Probability $P\left(\frac{H}{m_o^{1/2}} > x = \frac{H_o}{m_o^{1/2}}\right)$ of exceeding a normalized wave height $x = H_o/m_o^{1/2}$. The triangles are data, the solid line is the Rayleigh distribution (1) and the dashed line is the Weibull distribution (3) with $\alpha = 2.126$, $\beta = 8.42$. From Forristall (1978).

deep, may affect wave height statistics. Theoretical studies by Longuet-Higgins (1963) and laboratory experiments by Huang and Long (1979) have yet to be applied to wave height statistics.

A direct consequence of the failure of the Rayleigh distribution is that if, as is now the practice, the significant wave height is defined by (2), the correct factors relating $H_{sig}$ to lower probability waves are smaller than would be obtained from Rayleigh statistics. For example, given Forristall's (1978) empirical Weibull distribution, the expected maximum wave height in a 20 minute sample of waves of period $T_z$ = 12 sec (100 waves) would be $H_{100} = 5.917\ m_o^{1/2} = 1.48\ H_{sig}$, rather than $6.449\ m_o^{1/2} = 1.61\ H_{sig}$ under Rayleigh statistics. Extrapolating to a three hour period [~1000 waves], we find $H_{100} = 1.76\ H_{sig}$ empirically, rather than $1.936\ H_{sig}$ for Rayleigh-distributed waves. The expected maximum wave height in N waves for a distribution of the form (3) is obtained for large values of N from

$$H_{max} = m_o^{1/2}(\beta \ln N)^{1/\alpha}\ [1 + \gamma/\alpha \ln N], \qquad (4)$$

with $\gamma = 0.5772$ (Forristall, 1978), and is readily calculated for any fit to the data of the form (3).

The zero-crossing period $T_z$ is well correlated to the wave heights: that this should be so is not surprising, given the observation that the period of the peak of the wind wave spectrum increases with fetch, as the wave amplitudes grow. As noted earlier, Wilson and Baird (1972) found strong correlations between the zero-crossing period $T_z$ and the period of the spectral peak $T_p$. Observational evidence on the evolution of wind wave spectra (LeBlond & Mysak, 1978, p482ff; Mitsuyasu et al., 1980) shows that the total spectral energy increases linearly with the fetch while the period of the peak grows as the fetch to the one-third power. One then expects a good correlation between $H_{sig}^{1/2}$ and $T_z^{1/3}$. There is, however, a lot of scatter in the data; Scott (1965) gives the empirical relation

$$\frac{T_m}{2\pi} = 0.03\ H_{sig} + 1.35, \qquad (5a)$$

Seaconsult

where $T_m$ is the period of the peak of the spectrum (in seconds) and $H_{sig}$ is in feet. Arhan et al. (1979) on the other hand, give

$$0.110 < H_{max}/(T_{H_{max}})^2 < 0.149 \qquad (5b)$$

where the subscript max refers to the extreme wave height and its period, a result which is more directly relevant to the extreme waves but is subject to much scatter. The success of numerical wave models, as reported later, in reproducing observed wave heights as well as periods from knowledge of the winds, lend support to the existence of an intimate relation between H and $T_z$, even in the presence of considerable observational scatter.

## 3.2 Wind Wave Generation

A brief review of wind-wave generation models is in order as background for a discussion of hindcast design-wave estimation techniques.

All wave generation models require the windfield as input; they differ in the degree of sophistication with which they model the physics of wave growth and propagation. Models may be divided into two classes: parametric and spectral models. In parametric models, the sea state is specified in terms of parameters such as significant wave height and wave period as a function of wind fetch and duration. The classical model of this type is due to Sverdrup, Munk and Bretschneider (U.S. Army Corps of Engineers, 1973); more recent models are those of Hasselmann et al. (1976) and of Kawai et al. (1979). Spectral models describe the evolution of the sea state in terms of discrete frequency and direction bands, representing the evolution of the spectrum $E(\omega,\underset{\sim}{\kappa},t)$ in time. Various spectral models differ in their methods of specifying energy input functions and in their treatment of wave-wave interactions. The most complete spectral models such as that of Resio and Vincent (1977, 1978), include wave-wave interactions explicitly.

Resio and Vincent (1979) have compared sea-states generated by a selection of parametric and spectral models under the same wind conditions. They found that although differences exist between model predictions, these differences are most pronounced at short fetches and durations. Thus, for the worst sea-state conditions, the more modern parametric (Hasselmann) and spectral models (Resio and Vincent) agree closely.

In addition to feeding energy from the wind into the waves, a wind wave model must also propagate to the point of interest waves generated in surrounding areas. Thus, hindcast models include a propagation as well as a wave generation component. A useful assessment of the state of the art in hindcast modeling is provided by the results of Ewing et al. (1979), who used the hybrid parametric wind-generation and swell propagation model of Günther et al. (1979) for hindcasting waves for 42 storms in the North Sea. Comparison of hindcast waves with measurements of the maximum wave heights during 15 of these storms gave

$$H_{measured} = 0.985\ H_{model}$$

with a standard error of 0.7 m (10% of the mean). When all wave heights (not just the extremes) were included, the scatter reached 25% of the mean. Maximum wave periods are also satisfactorily predicted, with

$$T_z\ (measured) = 1.11\ T_z\ (model).$$

Comparisons of measured and modelled values of significant wave height and period are shown in Figure 3. Günther et al. (1979) feel that this is as good agreement as can be expected, given the limitations of the meteorological information and the inherent variability of the wave field. Gulf of Mexico hindcasts (Bea, 1975; Cardone et al., 1976) also produce adequate sea-states.

We may then conclude that existing wave generation theories are capable, when based on adequate wind information, of reproducing observed sea-states within satisfactory limits. Since most models give similar results, the limiting factor in the accuracy

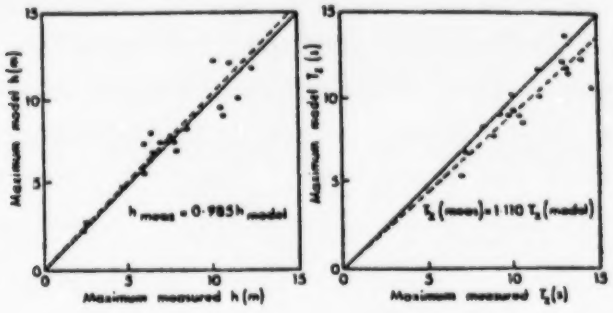


Figure 3 Comparison of maximum measured and model values of significant wave height and zero-crossing period at stations Famita and Stevenson. The dashed line shows the least-squares fit to the data; the solid line is the ideal perfect correlation line.

of wind wave models seems to be the description of the wind field itself. This description over the ocean is no simple task, since the surface pressure charts from which the wind field is to be calculated are drawn without many actual observation points over the sea and tend to be smoothed by the paucity of data. This is one of the most fundamental problems facing the practice of wave hindcasting. The surest way to inspire confidence in the results of hindcasting techniques in any given area is, of course, to compare hindcast wave characteristics with observed waves, as has been done by above-mentioned authors in the North Sea and the Gulf of Mexico. Because of different conditions of meteorological coverage on the open-sea coasts of Canada, it is not *a priori* evident that the same hindcast models would be as successful there as they have been in other areas and their credibility should be demonstrated. There is indeed evidence (R. Goodman, priv. comm.) that the wind data which can be calculated from surface pressure maps on file with FNWC (Fleet Numerical Western Centre, Monterey) completely missed storms which gave rise to an extreme storm surge ($D_o(T) \simeq 2.5$m) in the Beaufort Sea (Sept. 14-16, 1970) and to a lesser surge, confirmed by water level data, on Aug. 26-28, 1977.

## 3.3 Storm Surges

Residual sea-level variations (on the "long" time scale T) of the order of 4 m or more can be induced by wind and pressure forces in shallow waters. Given atmospheric inputs and the bottom topography, storm surges and associated currents are calculated by solving the long-wave shallow water equations. Reviews of the subject have been presented by Bretschneider (1967) and Reid et al. (1977); recent applications to the western Norwegian coast (Martinsen et al., 1979), to the southern North Sea (Dolata and Engel, 1979) and to the Louisiana Coast (Crawford, 1979) show that present models can reproduce the departures of water levels from predicted tidal values to within at best 10 cm. Coarse parameterisations of bottom roughness, limited wind resolution in space and time over the sea, and lack of information at open boundaries make higher accuracy improbable.

Although the most catastrophic consequences of storm surges occur on areas which are normally dry, their influence on displacing the sea level may also be extremely important in shallow offshore areas, causing flooding (of artificial islands) or bottoming out (of fixed platforms) in association with storm waves. Since it is not possible to predict storms in advance, probable storm surges at a given location should be inferred on the basis of past surges and from numerical calculations of the effects of past storms. Some steps towards estimating storm surge probabilities and frequencies in the North Sea have already been taken (Fuhrboter, 1979). Since the same data are required for wind-wave hindcasting as for storm surge calculations, both could be run in parallel for shallow areas. Storm surge hindcasting would produce one extreme value of displacement $D_o(T)$ for each storm modelled: these extreme values would then form the statistical data base for inferring probable extreme values of $D_o(T)$ during future storms, in much the same way as hindcast extreme wave heights are used, as described in Chapter 7.

## 4. Methods of Inference

The three basic methods described below are all subject to the same fundamental limitations of statistical inference. The observations of wave heights from which the present knowledge of sea-state are based are just a sample of values taken from a statistical population whose characterizing parameters are varying with time. Any estimate of these parameters is limited in accuracy by considerations of sampling accuracy and duration as well as by the form of the statistical distribution. The first step in any statistical inference program is then to anchor as firmly as possible the data base on which the inferences will be founded. This includes considerations of instrumental data collection, the gathering of ancillary data to determine if there exist distinguishable populations of the basic variable under different environmental conditions, and the choice of functional forms or of parameter values to characterize the wave statistics. Broad discussions of such considerations have been presented by many authors (Borgman, 1973; Tann, 1976; Borgman and Resio, 1980).

With information in hand of observed waves and/or of the winds that caused them, and with a proper appreciation of confidence limits associated with these data, one must make some assumptions about the future to infer anything about it from the present. Stationarity, i.e. the independence in time of broad environmental conditions, is normally assumed: this means that the future is similar to the past and that the probability distributions and properties of future storms as well as the statistical properties of the waves which they produce are those which apply now and which have been extracted from past data. To go beyond the stationarity hypothesis requires climatic modelling, based on extrapolation of recent climatic trends (as in Rye, 1976) or on some expectations of future storm characteristics.

The three methods discussed below are based respectively on 1) extrapolation from the tail of the observed probability

distribution, 2) the techniques of extreme value analysis and, 3) a combination of storm statistics and wave generation modelling to supplement the instrumental data base with hindcast wave data. In each case, we shall present the techniques used to obtain the design wave estimate, the assumptions on which the method is based, the information necessary for its application, and the reliability of the results.

## 5. Extrapolation from the Observed Distribution

The most straightforward technique to obtain an estimate of the expected extreme wave height $H_{max}$ in a long period of time, including say N waves, would be to apply the result of equation (4) directly. This is indeed what is done routinely to obtain a 3 hr-maximum wave height from a 12 minute (the sampling period used by the Institute of Ocean Sciences [U.K.]); Tann, 1976) or a 20 minute instrumental record (the MEDS sample duration). The number of waves in three hours is estimated by dividing that period of time by the wave period $T_z$ measured during the shorter sampling period: this gives N for substitution into (4). Naturally, this method provides a valid estimate of $H_{max}$ (3 hr) only if conditions during the instrumental recording period are truly typical of the three hour period following it: stationarity must be assumed over that time scale. Because sea-states may change appreciably over a three hour time scale, it has been suggested (Arhan, 1979) that continuous recordings be carried on during storm periods.

The direct extrapolation method is thus a basic step of further estimates based on 3-hourly maximum wave heights. Standard errors in estimates of $m_{0}$ and $T_z$ from the instrumental record are given by Cavanié (1979) as $\sigma(m_0)/m_0 \simeq 0.5/\sqrt{n}$ and $\sigma(T_z)/T_z \simeq 0.52/\sqrt{n}$ where n is the number of zero-crossings during the sampling period. For 12 sec waves in a 20 minute period, n = 100, and relative errors in the root of the variance and in the period $T_z$ are both about 5%. For $N = 3\ hr/T_z$ sufficiently large, we find from (4) that the relative error in the three-hourly expected maximum wave height resulting from errors in estimates during the sampling period is

$$\frac{\sigma(H_{max})}{H_{max}} \simeq \frac{\sigma(m_0)}{m_0} + \frac{1}{\alpha\beta \ln N} \frac{\sigma(N)}{N} \qquad (6)$$

Since $|\sigma(N)/N| = \sigma(T_z)/T_z$, and the distribution is nearly Rayleigh ($\alpha\beta \simeq 16$) the second term is negligible and the relative error in $H_{max}$ (3 hr) on account of sampling uncertainty is the same as that in estimating $m_0$. This source of error is discussed

here mainly because it affects the error levels in estimating the parameters specifying the extreme value distributions discussed in the next section. There are also, of course, confidence levels in the estimates of $\alpha$ and $\beta$ in the basic distribution to which the sample statistics are fitted.

Ochi (1973) noted that the most probable extreme value of N waves obeying the Rayleigh distribution, as given by

$$H_{max} = (8 \ln N)^{\frac{1}{2}} m_o^{\frac{1}{2}}, \qquad (7a)$$

(obtained from (4) with $\alpha = 2$, $\beta = 8$ and neglecting the second term in the parenthesis) has a probability of about 60% of being exceeded in any realization of N waves. Ochi then argued that a more useful estimate for engineering purposes would be that extreme value of H which has a small probability ($q << 1$) of being exceeded over a specified interval. This he showed to be equal to

$$H_{max}(q) = (8 \ln \frac{N}{q})^{\frac{1}{2}} m_o^{\frac{1}{2}} \qquad (7b)$$

for Rayleigh distributed waves. Extensions to the more general Weibull distribution are trivial: just replace N by N/q in (4). It is, of course, a straightforward mental leap to conclude that the wave height which is not exceeded more than q times in one sample will not be exceeded more than once in 1/q samples. This is the crucial inferential step on which are based all the methods described here: it depends critically on the assumption of long-term stationarity.

Ochi (1973) also showed that the result (7b) for the most probable extreme waves in 1/q realizations is the same as would be predicted by applying order statistics to the extreme wave heights of each realization: this method is then equivalent, for $q << 1$, to the extreme wave analysis discussed below.

The application is simple: given $m_o^{\frac{1}{2}}$ based on wave data over as many sampling periods as possible, including the more severe conditions encountered (which are very important in estimating

the variance; Ochi and Whalen, 1979), the most probable maximum wave height over 1/q sampling periods of N waves each is estimated from (7b). For example, for 12 sec waves over a 20 min sampling period, N = 100; over 20 years, there are about $5 \times 10^5$ 20 min periods and we find

$$H_{max}/m_o^{\frac{1}{2}} \simeq 12, \text{ i.e. } H_{max} \simeq 3\, H_{sig}.$$

This method relies heavily on stationarity and on sufficiency of the existing database to determine the wave height distribution accurately. The main problem with inferring extreme values over a period of many years by extrapolating the tail of a distribution determined over one (or even a great number) of short instrumental records is that the confidence intervals of the estimates diverge rapidly as the distribution is extrapolated beyond its supporting data. This point is common to all methods considered and is discussed at greater length in the next section; Figure 7 illustrates the point graphically. More specifically, if the probability that one wave of period $T_z$ has a height less than H is $[1 - P(H)]$, then the probability Q that all of N waves be less than H is $Q = [1 - P(H)]^N$ if all these waves are uncorrelated. N is a very large number if we are contemplating applying the statistical inference to a period of many years. If Q = 0.5, the probability is one-half, i.e. there is a one-half chance that all waves are below H; the upper quartile is Q = 0.75, the lower Q = 0.25. Using the fact that $N \gg 1$, so that $Q^{\frac{1}{N}} \simeq 1 + \frac{1}{N}\ln Q$ and $\ln(1-Q^{\frac{1}{N}}) \simeq \ln(-\frac{1}{N}\ln Q)$, we find for the basic Weibull distribution (3) that

$$\frac{H_o}{m_o^{\frac{1}{2}}}(Q) \simeq \beta^{\frac{1}{\alpha}}[-\ln\tfrac{1}{N}]^{\frac{1}{\alpha}}\left[1 + \frac{\ln(-\ln Q)}{\alpha \ln\frac{1}{N}} \ldots\right]. \qquad (8)$$

Since $\frac{\partial H_o(Q)}{\partial Q} > 0$, a higher percentile maximum wave estimate grows faster as N increases than a lower percentile estimate and the difference between $H_o(Q)$ at any two different percentiles increases with N: the confidence intervals diverge, as seen in Figure 7. It is thus clearly advantageous to base the statistical inference process on as long as possible a database in order to avoid extrapolation far beyond the supporting data.

## 6. Extreme Value Analysis

Since it is the extreme wave heights which are of interest to the designer, it is quite a natural step to choose as a basic statistical population only the extremes over each sampling period rather than all the waves in each record. There are also significant advantages, from the point of view of describing the distributions and extrapolating to future times, to this process of skimming the values of greatest interest from the top of each sample histogram.

A general result of probability theory (Fisher and Tippett, 1928) shows that the statistical distribution of sample maxima from independent and identically distributed (Rayleigh or otherwise) distributions must approach one of only three forms for large sample sizes: these are known and specified as

Fisher-Tippett Type I:

$$P(H<x) = \exp[-\exp-((x-A)/B)] \qquad (9a)$$

Type II:

$$P(H<x) = \begin{cases} 0 & x<A \\ \exp\left[-\left(\frac{x-A}{B}\right)^{-C}\right] & x\geq A \end{cases} \qquad (9b)$$

Type III:

$$P(H<x) = \begin{cases} \exp\left[-\left(\frac{A-x}{B}\right)^{C}\right] & x\leq A \\ 0 & x>A \end{cases} \qquad (9c)$$

with $B, C \geq 0$.

These distributions are known under a variety of names; we shall simply refer to them as FT-I, II, III. The question of choosing which of the three distributions is a better or more logical fit to extreme wave height statistics is not fully resolved. On the one hand, the types of distributions (such as [3]) which describe the statistics of wave heights in instrumental records fall into the "domain of attraction" (Pickands, 1975) of the

FT-I extreme value distribution: that is, the extreme values of waves so distributed would tend to obey (9a) without restrictions on the value of the spectral width parameter (Walden and Prescott, 1980). On the other, extreme wave heights are certainly bounded below, so that (9b), with A=0 would seem appropriate; again, there is also an upper bound in wave heights set by breaking, and a value of A dependent on water depth in (9c) could also be a logical choice.

All three types have been used in practice; Thom (1971) has used a FT-II distribution, bounded below, to describe extreme wave heights at a number of ocean weatherships. Suggestions (Borgman and Resio, 1980; Carter and Challenor, 1979) as to the relevance of the FT-III asymptotic distribution for wave heights bounded above have been made, but no specific applications have been found. Mainly because it fits the data as well as the other two distributions while requiring one less parameter to determine, the FT-I distribution has been used most frequently for fitting extreme wave height values.

Non-asymptotic types of distributions have also been used to fit the distribution of extremes; the Weibull distribution

$$P(H\leq x) = \begin{cases} 1 - \exp\left[-\left(\frac{x-A}{B}\right)^C\right] & x\geq A \\ 0 & x<A \end{cases} \qquad (10)$$

with $B, C \geq 0$ is used routinely in the design wave estimation used by the U.K. Institute of Oceanic Sciences (Tann, 1976).

## 6.1 Plotting Rules

Statistical inference based on extreme value statistics, then, consists first of finding as best a fit as possible of wave height values representing a homogeneous population of independent extremes to one of a choice of possible distributions. The fitted distribution can then be extrapolated to lower probabilities to estimate return periods or expected extreme values for

selected life times. The first problem is then one of curve fitting which may be reduced to a least squares linear regression by defining the "reduced variate"

$$r = \frac{(x-A)}{B} \tag{11}$$

which may be related to the observed probabilities $P(H<x)$.

For example, for FT-I, from (9a),

$$r = -\ln[-\ln P]. \tag{12a}$$

For a Weibull distribution (10),

$$r = [-\ln(1-P)]^{1/C} \tag{12b}$$

Starting from an ordered series, $X_1$, $X_2$, ...$X_i$, ...$X_N$ of extreme values, with $X_1 > X_2 \ldots > X_N$, probabilities $P(H<X_i)$ can be assigned according to a plotting position formula of the form

$$P(H<X_i) = 1 - \frac{(i-u)}{(N+v)} \tag{13}$$

where u and v are constants which depend on which distribution is to be fitted. We will come back to the choice of these constants later.

Having assigned a $P = P_i$ value to each $X_i$, $i = 1, \ldots N$, using (13) and calculated for each $P_i$ an $r_i$ value for the appropriate distribution being tested, one then plots $X_i$ vs $r_i$ to estimate the parameters and the goodness of fit of a line

$$X = A + rB$$

to the data $(X_i, r_i)$ with the usual least square regression technique or by other methods (Gringorten, 1963b).

Discussions of plotting methods and of the appropriate values of u and v to use in (13) have been presented by a number of authors (Blom, 1958; Kimball, 1960; Gringorten, 1963b; Draper, 1963). Borgman and Resio (1980) argue that the simple u = 0, v = 1 choice should be adequate for most purposes. A short table of values of (u, v) is given in Table I. Table II shows sample

Table I

Plotting factors u, v for a selection of distributions (from Petrauskas & Aagaard, 1971).

| | | u | v |
|---|---|---|---|
| Fisher-Tippett Type I: | (9a) | 0.44 | 0.12 |
| Weibull . . . . . | (10) | | |
| C = | 0.75 | 0.54 | 0.64 |
| | 0.85 | 0.51 | 0.59 |
| | 1.0 | 0.48 | 0.53 |
| | 1.1 | 0.46 | 0.50 |
| | 1.25 | 0.44 | 0.47 |
| | 1.5 | 0.42 | 0.42 |
| | 2.0 | 0.39 | 0.37 |

## Table II

**Fitting positions and calculated reduced variates for a set of extreme wave height values (from Petrauskas & Aagaard, 1971 and Borgman and Resio, 1980).**

| | | FT-I u = 0.44, v = 0.12 | | FT-I u = 0, v = 1 | | Weibull; C = 2.0 u = 0.39, v = 0.37 | |
|---|---|---|---|---|---|---|---|
| Rank: i | $X_i$ (feet) | $P_i$ | $r_i$ | $P_i$ | $r_i$ | $P_i$ | $r_i$ |
| 1 | 35.00 | 0.975 | 3.676 | 0.957 | 3.113 | 0.972 | 1.891 |
| 2 | 34.70 | 0.929 | 2.608 | 0.913 | 2.396 | 0.927 | 1.618 |
| 3 | 30.90 | 0.884 | 2.093 | 0.870 | 1.968 | 0.882 | 1.462 |
| 4 | 29.00 | 0.839 | 1.740 | 0.826 | 1.655 | 0.837 | 1.345 |
| 5 | 28.10 | 0.794 | 1.467 | 0.783 | 1.406 | 0.792 | 1.253 |
| 6 | 24.90 | 0.749 | 1.241 | 0.739 | 1.196 | 0.746 | 1.171 |
| 7 | 24.00 | 0.703 | 1.043 | 0.696 | 1.014 | 0.701 | 1.099 |
| 8 | 23.90 | 0.658 | 0.871 | 0.652 | 0.850 | 0.656 | 1.033 |
| 9 | 23.30 | 0.613 | 0.716 | 0.609 | 0.700 | 0.611 | 0.972 |
| 10 | 22.60 | 0.568 | 0.570 | 0.565 | 0.561 | 0.566 | 0.914 |
| 11 | 22.40 | 0.523 | 0.433 | 0.523 | 0.430 | 0.520 | 0.857 |
| 12 | 19.80 | 0.477 | 0.301 | 0.478 | 0.304 | 0.475 | 0.803 |
| 13 | 19.80 | 0.432 | 0.175 | 0.435 | 0.183 | 0.430 | 0.750 |
| 14 | 19.40 | 0.387 | 0.052 | 0.391 | 0.064 | 0.385 | 0.697 |
| 15 | 19.00 | 0.342 | -0.070 | 0.348 | -0.055 | 0.340 | 0.645 |
| 16 | 17.60 | 0.297 | -0.194 | 0.304 | -0.174 | 0.294 | 0.590 |
| 17 | 16.50 | 0.251 | -0.324 | 0.261 | -0.295 | 0.249 | 0.535 |
| 18 | 15.90 | 0.206 | -0.458 | 0.217 | -0.423 | 0.204 | 0.478 |
| 19 | 13.30 | 0.161 | -0.602 | 0.174 | -0.559 | 0.159 | 0.416 |
| 20 | 12.30 | 0.116 | -0.767 | 0.130 | -0.711 | 0.114 | 0.348 |
| 21 | 12.00 | 0.071 | -0.973 | 0.087 | -0.893 | 0.068 | 0.265 |
| 22 | 12.00 | 0.025 | -1.305 | 0.043 | -1.143 | 0.023 | 0.152 |

plotting calculations for a population of 22 extreme wave heights from a California West Coast location (Petrauskas and Aagaard, 1971) fitted by three different distributions; the fits are shown graphically in Figure 4. As noted by Petrauskas and Aagaard (1971), a typical sample of extreme wave data can be well fitted by a variety of distribution functions. The smaller the number and range of values ($X_i$, $r_i$) the more difficult it is to pick the "true" distribution with accuracy. Even if there is an underlying "true" distribution, its determination is difficult for small samples because of the natural scatter of data points about that distribution. An example, drawn from Petrauskas and Aagaard (1971) vividly illustrates the problem. Figure 5a shows two independent sets of random samples from the Weibull distribution with C = 2.0. Fits to each of these samples give widely diverging regression lines on each side of the "true" line. Another example, using synthetic data (Borgman & Resio, 1980), is shown in Figure 5b. The uncertainty associated with choosing the best distribution for the extreme values sample or, having chosen a distribution, of determining its parameter values, has important consequences on estimating a design wave height. There is no technique that can eliminate the natural variability of the data, and it is essential that this be borne in mind when arriving at numerical estimates.

## 6.2 Inferring a Design Wave Height

Having determined or chosen a distribution of extreme values $P(H < x)$, sampled on the average at an interval of $\tau$ years, we may assume under the simplest of circumstances that successive realizations of these values are uncorrelated, so that a value of H which can be exceeded with a probability q will be exceeded on the average once every 1/q trials. The return period for that value of H is then

$$T_R = \tau/q \qquad (14)$$

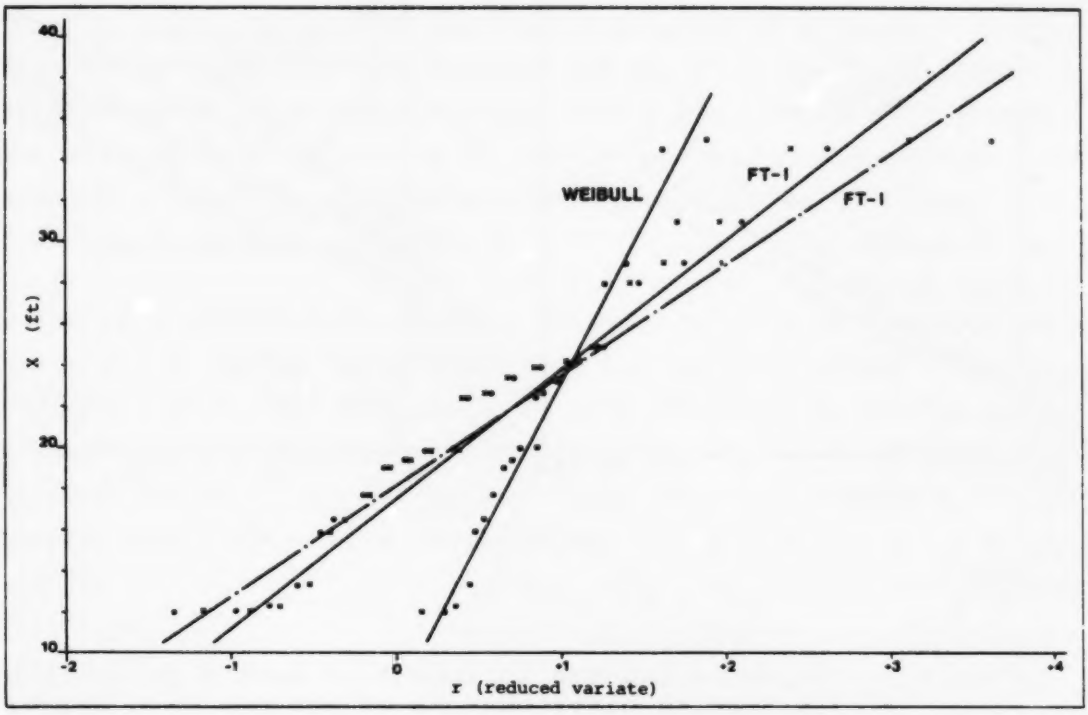


Figure 4 Fits of the lines X = Ar+B to extreme wave height data from Table II. Circles: plotting parameters u = 0.44, v = 0.12, fitted by FT-I curve —·—. Crosses: u = 0.0, v = 1.0, FT-I fit, solid line; Squares: u = 0.39, v = 0.37, fitted by a Weibull curve (equation 10) with C = 2.0.

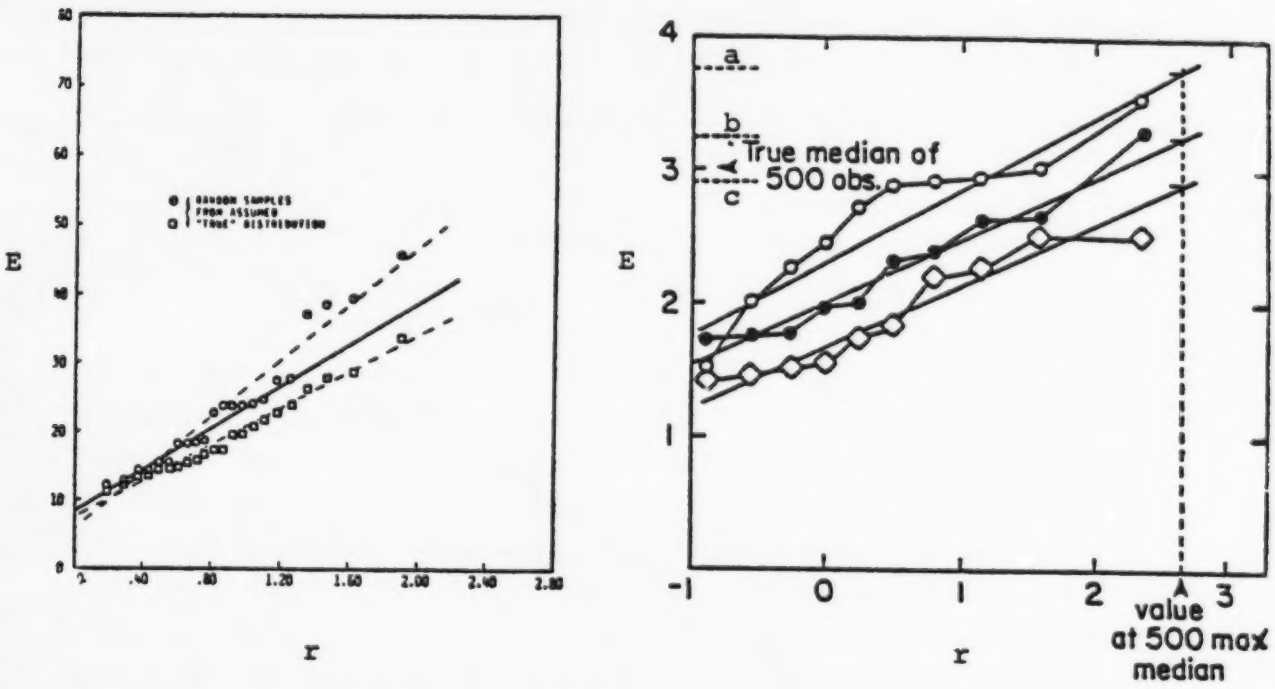


Figure 5. Illustrating the influence of sample variability in fitting distributions of maximum wave height E to the reduced variate r.

a) Two independent sets of random samples from a Weibull distribution. The solid line shows the assumed "true" relationship between extreme wave height E and the reduced variate r: $E = 15.2\ r + 8.19$. Two populations of random samples from the "true" distribution are shown by circles and squares respectively. Regressions from these sample populations diverge from the "true" fit and show the possible errors which may arise from a sparsely sampled data set. Example drawn from Petrauskas and Aagaard, 1971.

b) Three distributions of the maximum of 50 observations from a N(0,1) population (a Gaussian with zero mean and unit standard deviation), showing sample variability. Extrapolated median estimates from each sample distribution are indicated by the dotted horizontal lines a, b, c, compared to the true median for 500 observations, $E = 3$. From Borgman and Resio, 1980.

Since $q = 1 - P(H)$, we find the relation between return period and probability:

$$P(H) = 1 - \tau/T_R \qquad (15)$$

One may then calculate H from the appropriate distribution or read from a graph (such as Figure 4) the value of H which corresponds to P(H). If the extreme values are those obtained from a succession of 3 hour samples, $\tau$ = 3 hrs = $3.42 \times 10^{-4}$ yrs. For the data plotted in Figure 4, the 22 points correspond to extremes of a population of storm waves over 60 years: thus $\tau$ = 2.72 years/storm.

The 100-year expected extreme wave height thus has probability $P(H_{100}) = 1 - 2.72/100 = 0.9728$. This probability level corresponds, using (9a) and (10), to reduced variate values r = 3.59 and 1.89 respectively for the FT-I and the Weibull (C = 2.0) distributions. Values of $H_{100}$ read from Figure 4 curves then range from 37.3 ft to 40 ft.

## 6.3 Errors

As expected from the above discussion, errors in $H_{100}$ or any similar return-period wave height estimate can be attributed in part to uncertainty as to the appropriate distribution representing the distribution of extremes. Petrauskas and Aagaard (1971) have examined the effect of not knowing the "true" distribution; the expected maximum wave heights as a function of return period for the data of Table II are shown for a variety of distributions in Figure 6. All the distributions give similar results up to a 100-year return period, but diverge considerably for longer periods. We recall that the data for that particular estimate was based on 22 storms over 60 years, so that we would indeed expect coincidence over at least the sampling period. Another important source of uncertainty is based on sample variability, as illustrated in Figure 5 above. Estimates by Petrauskas and Aagaard (1971) indicate that the

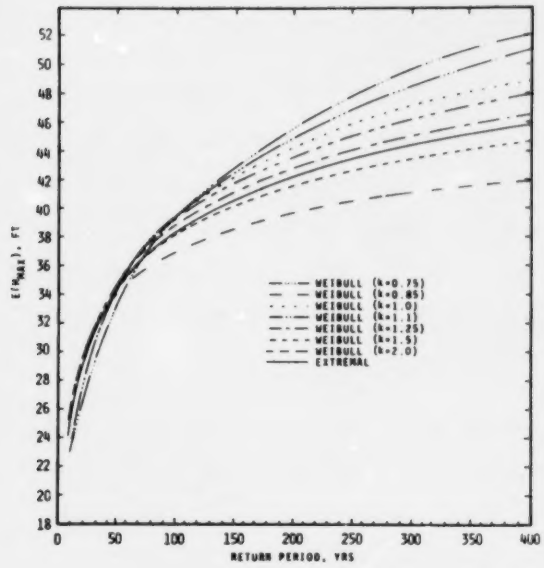


Figure 6 The effect on long-term extrapolation of not knowing the "true" distribution.
A comparison of a number of distributions fitted to the data of Table II. Note: the parameter k in the legend is what we have labelled C in equation (10). From Petrauskas and Aagaard (1971).

uncertainty based on sample variability is more serious than not knowing the "true" distribution. These authors as well as Gringorten (1963a) and Challenor (1979) give confidence limits for extreme value distributions to support the above. A more direct estimate of the range of variability about the expected extreme may be obtained following Borgman and Resio's (1980) argument, along lines already explored in Chapter 5.

The probability $P_i$ that any one sample has a height H smaller than a certain value $X_i$ is given in Table II; if all samples are independent, the probability Q that N successive samples are of a height $H<X_i$ is $(P_i)^N$. Thus

$$Q = P_i^N \qquad (16)$$

and the median (Q = 0.5) of the distribution of the N samples corresponds to the $0.5^{1/N}$ fractile of the original distribution. For the data of Table II and a 100 year period (N = 36.8 samples) the median 100-year extreme wave corresponds to the 0.981 fractile of the original distribution. Substituting in (12a) for an FT-I distribution, gives r = 3.97; other fractiles of the 100-year maximum are estimated in a similar fashion. The 0.25 and 0.75 fractiles (given by $0.25^{1/N} = 0.962$ and $0.75^{1/N} = 0.992$ fractiles of the original distribution) correspond to r = 0.325 and r = 4.85 respectively. For the FT-I curve of Figure 4 corresponding to the "x" points, these values of the reduced variate correspond to a median $H_{100}$ = 42 ft with 0.25 and 0.75 fractiles at 38 ft and 48 ft respectively. This range is indeed greater than the variability between the different curves.

## 6.4 Applications

Since the errors associated with the natural variability of the extrapolated wave height extremes exceed those due to uncertainty as to the form of the "true" distribution, there is little point in arguing in favour of any one of the distributions described by (9) and (10) as being the "best" for design wave prediction. Efforts should clearly be focussed on decreasing the uncertainty by extending the database in time so that,
i) the parameters A, B, C of the distribution chosen to represent the data can be determined as accurately as possible, and
ii) extrapolation for periods beyond the sampling duration is reduced as much as possible. Just as the various possible distributions which may be chosen to fit the data are seen (Figure 6) to diverge beyond the sampling period, the error

Seaconsult

bounds around any one distribution also spread out: an example is seen in Figure 7. A simple rule of thumb (Borgman & Resio, 1980) is not to extrapolate beyond twice the sampling period. Since design waves are often based on estimates of 50 to 100-year extreme values while instrumental data are rarely available over more than a few seasons, there is a clear need for a source of data extending over a longer time base. That is the very purpose of the hindcast method discussed next.

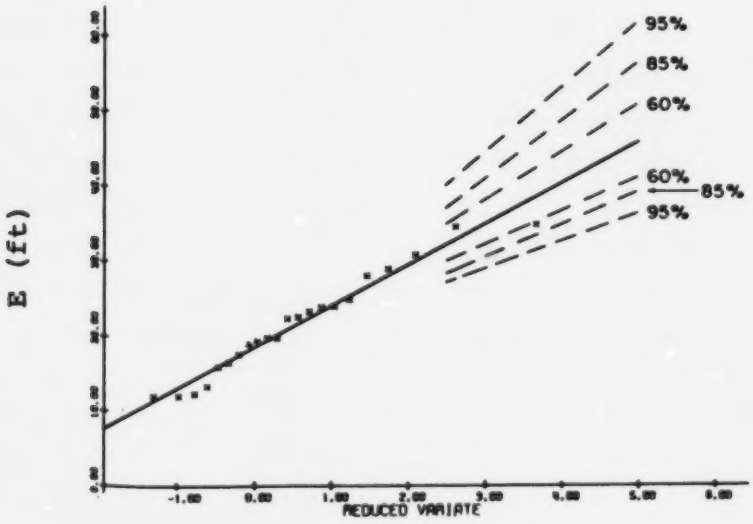


Figure 7 The spreading of confidence intervals about a FT-I fit to the extreme wave heights of Table II. From Petrauskas and Aagaard, 1971.

## 7. The Hindcasting Method

It is very often the case that although wave measurements may be available only for a few seasons, so that the direct sea-state statistics are rather meagre, meteorological data exist in the same area for a great number of years. Since it is possible to calculate the waves produced by a known wind field, as discussed earlier, it should then be feasible to supplement instrumental data with modelled waves calculated from historical storms. This method extends the database over many years, allows to concentrate on the strongest storms and promises a much more favourable basis for extrapolation into the future.

The basic premises of the hindcasting procedure are universally accepted: a wind-wave generation and propagation model using available meteorological parameters is used to calculate sea-states for all storms above a selected threshold level in the area of interest. That this hindcasting procedure can indeed reproduce observed sea states has been verified by the Ocean Data Gathering Program (Ward, 1974; Cardone et al., 1976) in the Gulf of Mexico, where a standard deviation of about 1 m was found between hindcast and observations, and by the North Sea Wave Model (Ewing et al., 1979) with errors of about 0.7 m. As pointed out by Haring and Heidemann (1980), it is not necessary that the hindcast model predict precisely each measured sea state. All that is required is that the hindcast model's aggregate statistics lead to the same distribution function for extremes as would measured data from the same storms.

Should there exist two or more distinct classes of storms (hurricanes, frontal systems), the waves produced by each class should be treated as belonging to a different population of random values so as to justify statistical inference based on population homogeneity (Borgman and Resio, 1980).

The process of statistical inference leading from hindcast waves to estimates of design wave and crest heights follows a series of steps; these have been discussed and explained by a number of authors (Petrauskas & Aagaard, 1971; Jahns & Wheeler, 1972, Haring & Heidemann, 1980) and are briefly reviewed as follows:

i) The highest wave in a sea-state

Given a probability distribution function $P_1(H<x)$ for the height of individual waves, related to wave intensity through the variance $m_o$, such as the Rayleigh (1) or the Weibull (3) distributions seen in Chapter 3, and an estimate of the zero-crossing period $T_z$, the probability $P_2(H;\sigma)$ that the maximum wave height in a sea-state $\sigma$ of chosen duration not exceed a certain value H, may be found as discussed in Chapter 5. The extreme wave height value for a sea-state will depend on the wave intensity during that sea-state, as specified by the spectral variance $m_o$ (obtained by the hindcast model) and on the number of waves in a sea-state (which depends on the duration of the latter and on the period $T_z$, also obtained from the hindcast model).

ii) The highest wave in a storm

A storm consists of a sequence of sea-states with gradually increasing wave heights up to a maximum value, followed by a decrease to lower levels. If successive sea-states are assumed independent of each other, the probability $P_3(H)$ that the maximum wave height during a whole storm not exceed a certain value is found from $P_3(H) = \prod_{i=1}^{M} P_2(H;\sigma_i)$, i.e. the combined probability over the M sea-states $\sigma_i$ of the storm that the specified value not be exceeded during any one of these sea-states. The consequences of there existing some degree of dependence between successive sea-states, as is the case for real storms, has been discussed by Tann (1976), who found that ignoring correlations between successive sea-states leads to errors of about 3% in extreme wave height estimates.

iii) Storm statistics

The most widely accepted statistical method is then to consider the N storms for which waves have been hindcasted as a random sample of a long-term distribution of extreme wave heights, and to fit this empirical distribution with an analytical function $P_4(H)$ such as the Fisher-Tippett or Weibull curves discussed in Chapter 6.

iv) Return periods

If storms occurred with clockwork regularity at fixed intervals, one could simply infer a return period from a scaling of the fitted distribution function, as per equation (15) in the previous section. Storm occurrences, however, also obey statistical rules and are well fitted by a Poisson distribution such that the probability of S storms occurring in a year is given by

$$P(S) = \lambda^S e^{-\lambda} / S!,$$

where $\lambda$ is the mean number of storms per year. The probabilities of occurrence of extreme wave heights must then be weighted by the storm probability. The probability that all wave heights in a random year are smaller than a certain level H will then be given by

$$P_5(H) = \exp[-\lambda(1-P_4(H))].$$

A return period $T_R$ (in years) can then be defined for a wave height H as

$$T_R = 1 / (1-P_5(H))$$

The probability that a certain wave height not be exceeded during the L years of a platform's lifetime is then simply

$$P_6(H) = [P_5(x)]^L.$$

The above sequence of calculations and estimations is the standard method of obtaining design wave values by the hindcast method. Simpler variants have also been suggested and applied.

A second, simpler method, advocated by Nolte (1973), does not represent the extreme storm waves as a random sample fitted by an analytical distribution function which may be extrapolated to find larger waves. Instead, the "random storm wave" is defined as having the average properties of the N hindcast storm waves. Thus in this case

$$P_4(H) = \frac{1}{N} \sum_{j=1}^{N} P_3(H_j ; \text{storm}_j)$$

i.e. the probability of a wave height H not being exceeded in a random storm is the average of the probabilities of this same wave height not being exceeded in the N hindcast storm. Since no extrapolation is possible, this method must include storms over a period of time at least as long as the life-time of structures of interest to produce reliable estimates.

Another simplified, method considers only the peak sea-state ($\sigma_{max}$) in each hindcast storm, ignoring the lower sea-states. Each storm thus contributes only one maximum significant wave height. An extreme wave height over that sea-state is then obtained by using a relation between significant wave height (or the variance $m_o$) and the maximum expected wave height based on an appropriate distribution function $P_1(H<x)$. Haring and Heidemann (1980) have criticized a version of this variant which uses the Rayleigh distribution for $P_1(H)$ and a theoretical relation between wave height and crest height as not accounting for the observed statistics of wave heights and crest heights. However, the simplification consists basically in choosing the extreme sea-state rather than keeping contributions from lower sea-states, which is a reasonable assumption (Ewing et al., 1979); more realistic wave height and crest height distributions can be used in conjunction with this model than those criticized by Haring and Heidemann (1980) in the example which they worked out.

A comparison of the three methods has been made by Haring and Heidemann (1980) for Gulf of Mexico hindcast waves resulting from 22 severe hurricanes over a 75 year period. Their conclusion is that the rare wave height estimates are not strongly influenced

by the method of statistical analysis employed: the 100-year extreme wave heights predicted by the three methods outlined in the foregoing were all within 70± 2 feet. As discussed in Chapter 6, the errors associated with sample variability are typically about twice as wide: i.e. ± 5 feet. The third method, using some perhaps objectionable approximations, was found to be rather more sensitive to the duration of the selected sea state duration than the others.

On the basis of the success of the wave hindcasting techniques in reproducing observed sea states, and in view of the lack of significant dependence of the estimates of design wave heights on the details of the statistical techniques used, one must conclude that the accuracy of the hindcast method in determining design wave values is not significantly limited by the wave generation models or by the choice of techniques of statistical inference. Rather, errors in the hindcast method will be caused 1- by possible errors in the description of the surface wind field, mainly due to the coarseness of data grids, and 2- by the natural sample variability in the extreme wave height population (illustrated in Figure 5) which affects the fit of a distribution function used in extrapolation.

## 8. Special Considerations

### 8.1 Wave Breaking

Some of the assumptions made in describing wave statistics no longer hold when waves begin to break. For example, whenever wave breaking is likely to occur frequently, the wave heights must be bounded above and distributions such as the Weibull or Rayleigh curves cannot describe actual wave height above some maximum value, depending on water depth. Haring et al. (1976) introduced depth-dependent corrections to the Rayleigh distribution to describe shallow water effects. There is also evidence (Haring & Heidemann, 1980) that the 100-year wave height decreases in shallower water (at least in the Gulf of Mexico), an effect which may be due to attenuation by breaking; similar remarks have been made by Bouws (1979) for waves in the southern North Sea. Monahan and Muircheartaigh (1980) estimate that up to about 20% of the sea surface is covered with breakers in a fully developed sea state under 40 knot winds: breaking must also play a role in limiting the wave height distribution in deep water and is perhaps responsible for the departures from the Rayleigh distribution. The lack of proper understanding of the role of breaking on wave statistics does not, however, seem to affect the possibility of achieving empirical descriptions of wave height distributions and is thus not an impediment to practical design wave estimation.

### 8.2 Local Effects

Since wave-producing storms occur over scales of hundreds of kilometres, the scale of resolution of design wave estimates of the hindcast wave model is much broader than that of some small scale features which can affect wave heights. The most obvious source of small scale variations is refraction by bottom topography over the continental shelf. Models of wave generation and propagation which correctly include bottom topography can, of course, handle these smaller scale variations,

at the cost however of having to decrease the grid size to represent the bathymetry well. Another source of local variability is due to refraction by current shears rather than by depth variations. This is, of course, much more difficult to describe since the zones of strong shear where this kind of effect can occur (the edge of the Gulf Stream or of the Labrador Current, frontal zones, large internal waves) vary in position and strength with time in a way which cannot be predicted with any reliability. Thus, although bottom refraction calculations could reasonably be expected to be performed on a site-specific basis to examine their possible contributions to the local wave climate, it is not reasonable with the information now available, to expect the same level of analysis for refraction effects by currents. Some examination of plausible worse-case effects and of their qualitative likelihood should however be attempted in regions of strong horizontal current shears.

## 9. Conclusions and Recommendations

The evolution of methods of estimating design-wave parameters has been guided by two main trends: 1) the need to strengthen the database in order to minimize the effects of sample variability, and 2) the desirability of extending the supporting database in time to avoid extrapolation far beyond the sampling interval. The succession of methods utilized illustrates this evolutionary trend: from an extrapolation of the tail of the distribution of individual wave heights, to an analysis of extreme values of a number of sea states, and finally to a method which includes data from all large historical storms.

It has been demonstrated that physical models used in wave hindcasting are capable of reproducing observed sea states with satisfactory accuracy. Because the hindcast method includes in its database all wind-generated sea-states above a selected threshold level over the longest available documented period, we conclude that it is the optimal method for arriving at a reliable estimate of design wave parameters under conditions where long (many decades) series of instrumental data are not available. A basic recommendation follows quite naturally:

i) It is recommended that a method based on wind-wave hindcast modelling be recognized as the most reliable means of estimating design wave parameters, given the relative length of meteorological vs. instrumental wave records.

As we have discussed above, there are many variations on the basic hindcasting method and it is clear that calibration and scientific judgment must be applied at many critical steps to avoid a number of pitfalls in its application. Our second recommendation, in its many parts, delineates the requirements for successful implementation of a hindcasting wind-wave model.

ii) The following steps are recommended as necessary for implementing the hindcast modelling method.

a) The basic meteorological information must be capable of resolving all storms above a certain threshold level and of distinguishing between different storm populations if such a distinction is appropriate (i.e. hurricanes versus frontal systems, for example). Failure to resolve small but intense storm systems may seriously affect the reliability of the design-wave estimates, and a discussion of the consequences of possible such failures should form part of the design wave analysis.

b) The record of storm frequency and intensity should be examined for the presence of significant climatic trends and the influence of any such trend taken into account into the probability distribution of storm statistics.

c) A physically correct hybrid wave-generation and propagation model should be used to relate winds to waves. Although parametric models have been shown to provide acceptable results in many cases, spectral models which include the details of energy transfer mechanisms are more successful in situations of rapid changes in wind direction and speed and should be recognized as providing a higher level of reliability, at the cost of more extensive computing. Regardless of which method is chosen for routine inplementation of the technique however, it should be tested by comparison with instrumental wave data for a sample of storms in each identified category.

d) Before any curve-fitting or statistical inference is contemplated, the instrumental record (if available) or the regional bathymetry and hydrography should be consulted for evidence of significant wave height limitations or enhancements by refractive processes, and for the influence of such effects on local wave height statistics as they pertain to the relation between spectral levels (as typified by the variance $m_o$) and significant or extreme wave heights.

e) Since sample variability has been seen to have a greater effect on design wave estimate than the precise form of the distribution function chosen to fit the sample of extreme wave heights, it is appropriate to choose the simplest curve, i.e. the Fisher-Tippett - Type I extreme value distribution to represent the hindcast wave data. The exception to this rule would be dictated by any evidence of the existence of an upper bound due to wave breaking, which would suggest a Fisher-Tippett Type III distribution; this kind of situation would be likely only in regions of very shallow water or strong wave-current interactions.

f) Estimates of storm surge extreme levels should be made in shallow water areas and added to estimates of wind-wave crest-heights for a proper estimation of the total maximum water depth to be expected at a specific site.

g) No estimate of extreme wave height or extreme sea-level displacement should be quoted without confidence intervals, indicating that there be a certain level of probability that the extreme wave height does not exceed certain values around the most likely estimate.

h) To permit performance of more extensive risk analyses, complete tested distribution functions rather than mere single-number estimates should be made available to users.

## Bibliography

Arhan, M.F., A.G. Cavanie and R.S. Ezraty, 1980. Determination of the period range associated to the design wave. Internat. Conf. Sea Climatology, Paris, 3-4 Oct. 1979; 7-26. Technip, Paris.

Arhan, M.F., 1979. Analysis of the variability of spectral estimates using long continuous wave records. Internat. Conf. Sea Climatology, Paris, 3-4 Oct. 1979; 27-49, Technip, Paris.

Bea, R.G., 1975. Gulf of Mexico hurricane wave heights. J. Petroleum Technology, (Sept. 1975), 1160-1172.

Blom, G., 1958. Statistical Estimates and Transformed Beta Variables. Wiley & Sons, N.Y.

Borgman, L.E. and D.T. Resio, 1980. Extremal statistics in wave climatology. Unpublished manuscript, 45 pp.

Borgman, L.E., 1973. Probabilities of highest wave in a hurricane. J. Waterways, Harbors and Coastal Eng. Div., ASCE, WW2, 185-207.

Borgman, L.E., 1963. Risk criteria. J. Waterways, Harbors and Coastal Eng. Div., ASCE WW3, 1-34.

Borgman, L.E., 1961. The frequency distribution of near extremes. J. Geophys. Res., 66, 3295-3307

Bouws, E., 1980. Spectra of extreme wave conditions in the southern North Sea considering the influence of water depth. Internat. Conf. Sea Climatology, Paris, 3-4 Oct. 1979; 51-71. Technip, Paris.

Bretschneider, C.L., 1967. Storm surges. Advances in Hydroscience, 4, Academic Press, 341-418.

Cardone, V.J., W.J. Pierson, and E.G. Ward, 1976. Hindcasting the directional spectra of hurricane-generated waves. J. Petrol. Tech., (April 1976), 385-394.

Carter, D.J.T. and P.G. Challenor, 1980. Return wave heights estimated from monthly maximum values. Internat. Conf. Sea Climatology, Paris, 3-4 Oct. 1979; 271-294. Technip, Paris.

Cartwright, D.E. and M.S. Longuet-Higgins, 1956. The statistical distribution of the maxima of a random function. Proc. Roy. Soc., A, 237, 212-232.

Cavanié, A.G., 1980. Evaluation of the standard error in the estimation of mean and significant wave heights as well as mean period from records of finite length. Internat. Conf. Sea Climatology, Paris, 3-4 Oct., 1979; 73-88. Technip, Paris.

Challenor, P.G., 1979. Confidence limits for extreme value statistics. Inst. Oceanogr. Sciences, Rep. No. 82, 7 pp. + figs.

Crawford, K.C., 1979. Hurricane surge potentials over south-east Louisiana as revealed by a storm-surge forecast model: a preliminary study. Bull. Am. Met. Soc., 60, 422-429.

Dolata, L.F. and M. Engle, 1979. Sturmflut vorhersagen mit mathematisch-physikalischen Modellen. Die Küste, 34, 203-225.

Draper, L., 1963. Derivation of a design wave from instrumental records of sea waves. Proc. Inst. Civil Eng., 26, 291-304, Paper no. 6690.

Earle, M.D., 1975. Extreme wave conditions during Hurricane Camille. J. Geophys. Res., 80, 377-379.

Ewing, J.A., 1980. Numerical wave models and their use in hindcasting wave climate and extreme value wave heights. Internat. Conf. Sea Climatology, Paris, 3-4 Oct. 1979; 159-182. Technip, Paris.

Ewing, J.A., T.J. Weare and B.A. Worthington, 1979. A hindcast study of extreme wave conditions in the North Sea. J. Geophys. Res., 84, 5739-5747.

Fisher, R.A. and L.H.C. Tippett, 1928. Limiting forms of the frequency distribution of the largest or smallest number of a sample. Proc. Camb. Phil. Soc., 24, 180-190.

Foreman, M.G.G., 1977. Manual for tidal heights analysis and prediction. Pac. Mar. Sci. Rept. 77-10, Inst. Ocean Sciences, Sidney, B.C., 97 pp.

Forristall, G.Z., 1978. On the statistical distribution of wave heights in a storm. J. Geophys. Res., 83, 2353-2358.

Forristall, G.Z., E.G. Ward, V.J. Cardone and L.E. Borgmann, 1978. The directional spectra and kinematics of surface gravity waves in tropical storm Delia. J. Phys. Oceanogr., 8, 888-909.

Fuhrboter, A., 1979. Wahrscheinlichkeiten und Haufigkeiten von Extremsturm fluten. Die Küste, 34, 40-52.

Goodknight, R.C. and T.L. Russell, 1963. Investigation of the statistics of wave heights. J. Waterways, Harbors and Coastal Eng. Div., ASCE, WW2, 29-54.

Gringorten, I.I., 1963a. A simplified method of estimating extreme values from data samples. J. Appl. Meteorol. 2, 82-89.

Gringorten, I.I., 1963b. A plotting rule for extreme probability paper. J. Geophys. Res., 68, 813-814.

Gringorten, I.I., 1963c. Envelopes for ordered observations applied to meteorological extremes. J. Geophys. Res., 68, 815-826.

Gunther, H. W. Rosenthal, T.J. Weare, B.A. Worthington, K. Hasselmann and J.A. Ewing, 1979. A hybrid parametrical wave prediction model. J. Geophys. Res., 84, 5727-5738.

Haring, R.E., 1980. Extreme wave statistics based on NORSWAM hindcast results. Internat. Conf. Sea Climatology, Paris, 3-4 Oct., 1979; 313-361. Technip, Paris.

Haring, R.E. and J.C. Heideman, 1980. Gulf of Mexico rare wave return periods. J. Petroleum Technology, Jan., 1980, 35-47.

Haring, R.E., A.R. Osborne and L.P. Spencer, 1976. Extreme wave parameters based on continental shelf storm wave records. ASCE 15th Conference on Coastal Eng., 151-170.

Hasselmann, K., D. Ross, P. Muller and W. Sell, 1976. A parametric wave prediction model. J. Phys. Oceanogr., 6, 200-228.

Huang, N.E. and S.R. Long, 1979. A study of the surface elevation probability distribution and statistics of wind-generated waves. Unpublished Manuscript, 19 pp. + figs.

Jahns, H.O. and J.D. Wheeler, 1972. Long-term wave probabilities based on hindcasting of severe storms. Offshore Technology Conference, Paper OTC 1590, Houston, Tex. 1972.

Kawai, S., P.S. Joseph and Y. Toda, 1979. Prediction of ocean waves based on the single-parameter growth equation of wind waves. J. Oceanogr. Soc. Japan, 35, 151-167.

Kimball, B.F., 1960. On the choice of plotting positions on probability paper. J. Amer. Statist. Assoc., 55, 546-560.

Lawless, J.F., 1974. Approximations to confidence intervals for parameters in the extreme value and Weibull distributions. Biometrika, 61, 123-129.

LeBlond, P.H. and L.A. Mysak, 1978. Waves in the Ocean, Elsevier, 602 pp.

Long, R.B., 1980. The statistical evaluation of directional spectrum estimates derived from pitch/roll buoy data. J. Phys. Oceanogr., 10, 944-952

Longuet-Higgins, M.S., 1975. On the joint distribution of the periods and amplitudes of sea waves. J. Geophys. Res., 80, 2688-

Longuet-Higgins, M.S., 1973. On the form of the highest progressive and standing waves in deep water. Proc. R. Soc. Land. A, 331,

Longuet-Higgins, M.S., 1969. On wave breaking and the equilibrium spectrum of wind-generating waves. Proc. Roy. Soc. A, 310, 151-159.

Longuet-Higgins, M.S., 1963. The effect of non-linearities on statistical distributions in the theory of sea waves. J. Fluid Mech., 17, 459-480.

Longuet-Higgins, M.S., 1952. On the statistical distribution of the heights of sea waves. J. Marine Res., 11, 245-266.

Martinsen, E.A., B. Gjevik and L.P. Roed, 1979. A numerical model for long barotropic waves and storm surges along the western coast of Norway. J. Phys. Oceanogr., 9, 1126-1138.

Matushevskiy, G.V., 1977. Determination of maximum wind-wave heights in a storm. Oceanology, 17, No. 1, 14-16.

Mitsuyasu, H., F. Tasai, T. Suhara, S. Mizuno, M. Ohkusu, T. Honda and K. Rikiishi, 1980. Observation of the power spectrum of ocean waves using a cloverleaf buoy. J. Phys. Oceanogr., 10, 286-296.

Monahan, E.C. and I.O. Muircheartaigh, 1980. Optional power-law description of oceanic whitecap coverage dependence on wind speed. J. Phys. Oceanogr., 10, 2094-2099.

Murty, T.S., 1977. Seismic sea-waves: tsunamis. Dept. Fish. & Marine Services, Bull. No. 198, Ottawa.

Nolte, K.G., 1979. Joint probability of wave period and height. J. Waterway, Port, Coastal and Ocean Div., ASCE, WW4, 470-474.

Nolte, K.G. and F.H. Hsu, 1979. Statistics of larger waves in a sea state. J. Waterway, Port, Coastal and Ocean Division, ASCE, WW4, 389-404.

Nolte, K.G., 1973. Statistical methods for determining extreme sea states. Proc. 2nd Internat. Conf. Ports and Oceans 'Engineering under Arctic Conditions,' U. of Iceland, Aug. 1973, 1-38.

Ochi, M.K. and J.E. Whalen, 1979. Estimation of extreme waves critical for the safety of offshore structures. Offshore Technology Conference, Houston, Tex., paper OTC 3596.

Ochi, M.K., 1973. On prediction of extreme values. J. Ship Research, (March 73), 29-37.

Petrauskas, C. and P.M. Aagaard, 1971. Extrapolation of historical storm data for estimating design-wave heights. J. Soc. Petroleum Engineers, (March 1971), 23-37.

Pickands, J. III, 1975. Statistical inference using extreme order statistics. Ann. Statistics, 3, 119-131.

Reid, R.O., A.C. Vastano, R.E. Whitaker and J.J. Wanstratu, 1977. Experiments in storm surge simulation. The Sea, 6, Wiley-Interscience, 145-168.

Resio, D.T. and C.L. Vincent, 1979. A comparison of various numerical wave prediction techniques. 11th annual Offshore Tech., Calif., Houston, Tex., Paper OTC 3642.

Resio, D.T. and C.L. Vincent, 1978. A numerical hindcast model for wave spectra on water bodies with irregular shoreline geometry. Report 2: Model verification with observed wave data. Misc.Paper H-77-9. Hydraulics Laboratory, U.S. Army Engineer.Waterways Experiment Station, Vicksburg, Miss.

Resio, D.T. and C.L. Vincent, 1977. A numerical hindcast model for wave spectra on water bodies with irregular shoreline geometry. Report 1: Test of nondimensional growth rates. Misc. Paper H-77-9. Hydraulics Laboratory, U.S. Army Engineer.Waterways Experiment Station, Vicksburg, Miss.

Rye, H., 1976. Long term changes in the North Sea wave climate and their importance for the extreme value predictions. Marine Sci. Comm., 2, 419-448.

Rye, H., 1974. Wave group formation among storm waves. Proc. 14th Coastal Eng. Cont., Copenhagen, 1974. ASCE, 164-183.

Scott, J.R., 1965. A sea spectrum for model tests and long-term ship prediction. J. Ship. Res., 9, 142-152.

Tann, H.M., 1976. The estimation of wave parameters for the design of offshore structures. A description of the method presently used by IOS. Inst. Oceanogr. Sciences, Rep. No. 23, 29 pp.

Thom, H.C.S., 1971. Asymptotic extreme-value distributions of wave heights in the open ocean. J. Marine Res., 29, 19-27.

U.S. Army Corps of Engineers, 1973. Shore Protection Manual, Vol. I, Fort Belvoir, Va.

Walden, A.T. and P. Prescott, 1980. The asymptotic distribution of the maximum of N wave crest heights for any value of the spectral width parameter. J. Geophys. Res., 85, 1905-1909.

Ward, E.G., 1974. Ocean Data Gathering Program - An Overview. Sixth Annual Offshore Technology Conference, Houston, May 6-8, 1974. Paper OTC 2108-B.

Wilson, J.R. and W.F. Baird, 1972. A discussion of some measured wave data. Proc. 13th Coastal Eng. Conf., Vancouver, 1972. ASCE.